学校编码：**10384** 分类号

密级

学号：**21620060153302** **UDC**

厦门大学

博 士 学 位 论 文

深海沉积物来源宏基因组文库的构建与生物活性物质的功能筛选

Functinal screening of bioactive substance from metagenomic libraries constructed with deep sea sediments

汤熙翔

指导教师姓名：肖 湘 教授

专 业 名 称：微 生 物 学

论文提交日期：2010 年 6 月 28 日

论文答辩时间：2010 年 9 月2 日

学位授予日期： 年 月 日

答辩委员会主席： 教授

评 阅 人：

2010 年 6 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。

本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外，该学位论文为(　　)课题(组)的研究成果,获得(　　)课题(组)经费或实验室的资助，在(　　)实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名)：

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》

等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

(　　)1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

(　　)2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年　　月　　日

目录

摘 要 I

Abstract II

1.前言 1

1.1 宏基因组学的发展历程 1
1.2 宏基因组学在生物活性物质发现中的应用 8
1.2.1 功能筛选 10
1.2.2 序列同源性筛选 13
1.3 本文研究的思路、目的和意义 15

2.材料与方法 16

2.1 主要材料 16
2.1.1 深海沉积物样品 16
2.1.2 菌株与细胞株 16
2.1.3 常用培养基 16
2.1.3 主要试剂 17
2.1.4 主要仪器设备 17
2.2 主要方法 18
2.1.1 深海沉积物样品的富集 18
2.2.2 深海沉积物富集样品 DNA 提取 18
2.2.3 深海沉积物富集样品宏基因组文库的构建 19
2.2.4 宏基因组文库克隆子发酵产物萃取 21
2.2.4.1 克隆子生长曲线测定 22
2.2.4.2 文库克隆子发酵物提取 22
2.2.5 文库克隆子发酵浸膏 HPLC 分析 22
2.2.6 文库克隆子发酵浸膏质谱分析 22
2.2.7 文库克隆子发酵浸膏细胞毒活性筛选 22
2.2.8 文库克隆子发酵浸膏抗菌活性筛选 23
2.2.9 文库克隆子发酵浸膏抗病毒活性筛选 23
2.2.10 活性克隆子大规模发酵及发酵浸膏处理 24
2.2.11 克隆子转座突变分析 24
2.2.12 克隆子产尿黑酸 HPLC 验证 25
2.2.13 芳香族降解功能检测 25

3.结果与分析 26

3.1 深海宏基因组文库构建 26
3.2 深海宏基因组文库克隆子发酵产物化学多样性分析 27
3.2.1 文库克隆子生长曲线测定 27
3.2.2 文库克隆子发酵物提取 28
3.2.3 文库克隆子发酵浸膏 HPLC 结果分析 29

3.2.4 文库克隆子发酵浸膏质谱结果分析.....30
3.3 文库克隆子发酵浸膏抗肿瘤活性筛选.....32
3.4 文库克隆子发酵浸膏抗菌活性筛选.....49
3.5 文库克隆子发酵浸膏抗病毒活性筛选.....49
3.6 深海富集宏基因组文库产颜色克隆子分析.....51
3.6.1 深海富集宏基因组库产颜色克隆子的发现.....51
3.6.2 深海宏基因组库产颜色克隆子生物活性筛选及 HPLC 指纹分析.....52
3.6.4 深海宏基因组库克隆子 5C11 和 11F6 细胞毒活性化合物的分离鉴定.....67
3.6.5 产黑色素克隆子的功能基因分析.....80
3.6.6 克隆子中芳香族降解基因簇的降解功能初步分析.....85
4.讨论与分析.....90
4.1 深海宏基因组文库生物活性分子筛选的几个关键因素.....90
4.1.1 宏基因组 DNA 样品的制备.....90
4.1.2 合适的载体-宿主系统.....91
4.1.3 高通量、有效的筛选体系.....91
4.2 芳香降解基因簇与深海宏基因组文库来源生物活性分子的关系.....92
5.结论与展望.....94
5.1 结论.....94
5.2 展望.....94
参考文献.....96
附录.....107
表 1. 太平洋中国多金属结核区 Cosmid 文库混合克隆发酵粗提物.....107
表 2. Mothra 热液场 Fosmid 文库混合克隆粗提物.....112
表 3.10 个全长测序宏基因组文库克隆子基因注释.....127
致谢.....160

Contents

Chinese abstact ..... **I**

English abstract ..... **II**

1.Introduction **1**

1.1 History of the metagenomic technology 1
1.2 Progress on application of metagenomic technology in bioactive molecules identification 8
1.2.1 Functional screening 10
1.2.2 Sequence homology screening 13
1.3 Purpose,signification and contents of this thesis 15

2. Materials and methods **16**

2.1Main materials 16
2.1.1 Deep-sea sediment samples 16
2.1.2 Strains and Cell lines 16
2.1.3 Cultrue medium 16
2.1.3 Agents 17
2.1.4 Instruments 17
2.2Methods 18
2.1.1 Enrichment of deep-sea sediment 18
2.2.2 DNA extract of enrichment deep-sea samples 18
2.2.3 Construction of metagonomic libraries through pre-enriched methods 19
2.2.4 Fermentation extract of metagenomic library clones 21
2.2.4.1 Growth curve detemination of clones 22
2.2.4.2 Fermentation extracts of clones from libraries 22
2.2.5 HPLC analysis of fermentation extracts from metagenomic clones 22
2.2.6 MS analysis of fermentation extracts from metagenomic clones 22
2.2.7 Cytotoxic screening of fermentation extracts from metagenomic clones 22
2.2.8 Antibacterial screening of fermentation extracts from metagenomic clones 23
2.2.9 Antiviral screening of fermentation extracts from metagenomic clones 23
2.2.10 Large scale fermentation and extraction of bioactive metagenomic clones 24
2.2.11Transposon mutation analysis of metagenomic clones 24
2.2.12 HPlC confirmation of HGA 25
2.2.13 Detection of aromatic compounds degradation 25

3. Results and discussion **26**

3.1 Construction of metagenomic libraries with deep sea sediments 26
3.2 Chemical diversity analysis of fermentation products from deep-sea metagenomic libraries clones 27
3.2.1 Growth curve detemination of clones 27
3.2.2 Fermentation extracts of clones from libraries 28

3.2.3 HPLC analysis results of fermentation extracts from metagenomic clones.....29
3.2.4 MS analysis results of fermentation extracts from metagenomic clones 30
3.3 Cytotoxic screening of fermentation extracts from metagenomic clones 32
3.4 Antibacterial screening of fermentation extracts from metagenomic clones 49
3.5 Antiviral screening of fermentation extracts from metagenomic clones.....49
3.6 Analysis of color-producing clones from pre-enrichment metagenomic libraries 51
3.6.1Identification of color-producing clones from pre-enrichment metagenomic libraries.....51
3.6.2Bioactive screeing and HPLC analysis of color-producing clones from pre-enrichment metagenomic libraries.....52
3.6.4 Cytotoxic compounds isolation and chractrization of clones 5C11and 11F6.....67
3.6.5 Functional genes/gene clusters analysis of melanin-producing clones 80
3.6.6 Preliminary analysis of aromatic compounds degradation function of metagenomic clones containning aromatic compounds degradation gene clusters.....85

4. Discussion 90

4.1Some key points of bioactive screening from deep-sea metagenomic libraries.....90
4.1.1 Preparation of DNA for metagenomic libraries construction.....90
4.1.2 Suitable vector-host system.....91
4.1.3 High throughput and efficient screening system.....91
4.2 Relationship between aromatic compounds degradation gene clusters and bioactive molecules form deep-sea metagenomic libraries 92

5. Conclusions and perspectives..... 94

5.1 Conclusions.....94
5.2 perspectives.....94

References 96

Appendix 108

Acknowledgments..... 161

摘 要

深海占地球表面积的49%，是地球表面最大的生物栖息场所，深海环境中（微）生物基因资源的开发具有现实、重要的资源及科学价值。典型深海特征为低温(<5℃) 和高压 (最高达 110 MPa)，据估计在此极端条件下，深海环境中可培养微生物还不到0.01%，采用宏基因组技术能部分克服培养方法产生的偏差，获得新的科学发现和基因资源。

本论文采用富集方法构建了3个深海沉积物来源宏基因组文库，并对3个深海沉积物富集宏基因组文库及课题组前期构建的两个深海环境宏基因组文库克隆子发酵代谢产物进行了抗肿瘤（细胞毒）、抗菌、抗病毒活性的功能性筛选。

文库克隆子的化学多样性分析表明深海环境宏基因组文库克隆子发酵产物以大肠杆菌本底为主。对两个深海环境宏基因组文库克隆子发酵产物的生物活性筛选发现了几个具有抗肿瘤（细胞毒）活性的混合克隆和三个具有一定抗流感病毒活性的单克隆。

在 3 个深海沉积物富集宏基因组文库中发现了 16 个产颜色的克隆子，这些克隆子发酵产物具有抗真菌和/或抗肿瘤(细胞毒)活性。从其中两个克隆子 5C11 和 11F6 中分离鉴定出 6 个环二肽化合物和三个吲哚类化合物，其中一个为新结构的(2-羟基苯基)(二-1*H*-吲哚-3-基)乙酸，该化合物具有良好的细胞毒活性，对三种肿瘤细胞株的 IC_{50} 约 25 μg/ml。

通过对其中 10 个产颜色克隆子的全长测序，发现了一个与芳烃降解相关的功能基因簇及两个与酚类降解相关的功能基因簇。转座子突变及 HPLC 指纹图谱和细胞毒活性测定结果表明，这些功能基因簇与克隆子次生代谢产物的细胞毒活性密切相关。克隆子全长序列分析及转座子突变结果发现了两个与黑色素产生密切相关的 4HPPD 酶，一个尿黑酸代谢途径完整功能基因簇及一个与色氨酸代谢相关的单加氧酶。

我们的研究提示对深海样品进行特定类群微生物预富集的方法具有一定的方法学意义，可显著提高生物活性克隆子的筛选阳性率，从目前的小于$1/10^5$文库克隆子提高到大于$1/10^3$文库克隆子。

关键词：深海，宏基因组，生物活性分子，功能基因（簇）

Abstract

Deep-sea accounts for 49% of the earth's surface so as to be the earth's largest bio-habitat areas. Biomining the biological(microbial) genetic resources shows the great practical and scientific importance. Typical deep-sea environment is characterized by low temperature(<5℃) and high pressure(up to 110MPa). It is estimated that there may be less than 0.01% cultured microbes under this extreme circumstances. Metagenomic approach can be used to overcome at least partially the bias, gain access to new scientific discoveries and genetic resources.

Three metagenomic libraries with deep-sea sediments samples throμgh previously enrichment methods were constructed. Functional screening of antitumor(cytotoxic),antibacterial and antivirus activity from fermentation products originated from three enriched metagenomic libraries and two previously constructed environmental deep-sea metagenomic libraries clones was performed.

Chemical diversity analysis showed that fermentation products from deep-sea environmental metagenomic libraries clones were majorly from *E.coli* background. Functional screening resulted in indentification several antitumor (cytotoxic) mixed clones and three antiviral monoclones.

16 color-producing clones which showed antifungus and or antitumor(cytotoxic) were found in three enriched deep-sea metagenomic libraries. Six cyclo dipeptides and three indol derivatives compounds including a new structure and cytotoxic(IC_{50} aroud 25μg/ml) (2-hydroxyphenyl)(di-1*H*-indol-3-yl)acetic acid were isolated and characterized from two clones 5C11 and 11F6.

Throμgh full length sequencing of ten color-producing clones, one aromatic hydrocarbon degradation gene cluster and two phenol degradation gene clusters were identified. Transposon mutation analysis , HPLC fingerprints analysis and cytotoxic detection results demonstrated that these clusters are closely related with cytotoxic activity of second metabolism products from these clones. Full length sequencing and transposon mutation analysis resulted in identification of two 4HHP enzymes which closed related to melanin biosynthesize, one homogentisic acid metabolism cluster

and one flavin-binding monooxygenase related to tryptophan metabolism.

Our results indicated that pre-enrichment methods show great potential to improve positive probability of screening bioactive substance from deep-sea metagenomic libraries, the positive rate was up to less that $1/10^3$ clones from more than $1/10^5$ clones.

Key words: deep-sea, metagenome, bioactive molecule, functional gene(cluster)

1.前言

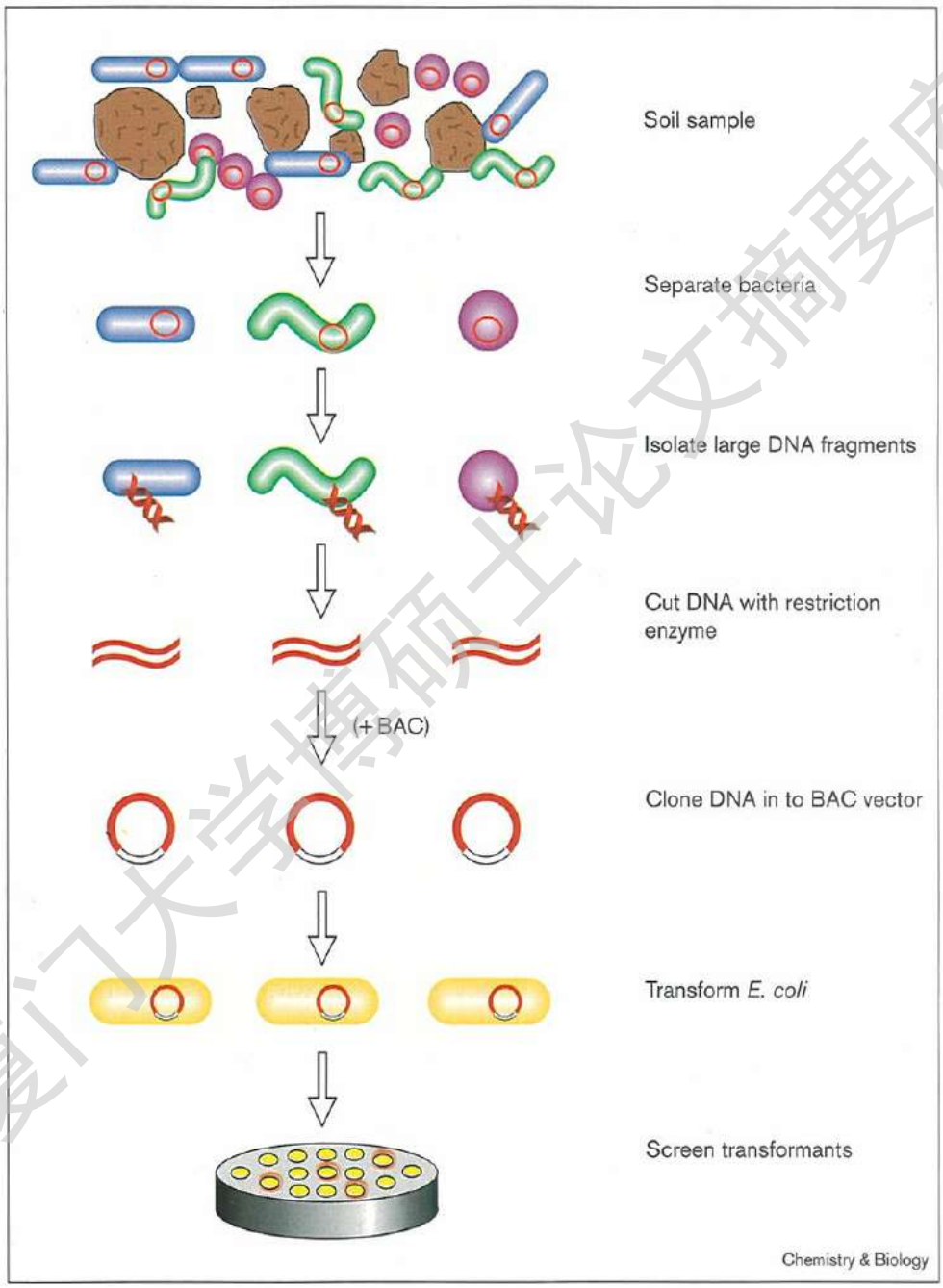
1.1 宏基因组学的发展历程

在生物圈中,微生物占据着重要的地位，贡献了地球生物多样性的2/3，其中深海环境所蕴藏的微生物占整个地球生物量的1/10-2/3，是未来新物种、新基因的主要来源。在距今35亿年前的古化石中就发现了微生物。据估计，地球上原核细胞的总量达到$4-6\times10^{30}$[1],包括10^{6}到10^{9}中独立的基因型[2]，如此复杂多样的微生物生态为利用微生物发现新的基因、基因簇及其表达产物提供了广泛的基础。然而由于微生物群落及其栖息环境的特殊性和复杂性,在实验室条件下,由于实验条件的选择性压力，难以模拟和重现其生活的原始环境条件,致使环境中仍有高达99 %的微生物未能得到纯培养,海水、淡水、中营养湖水、未污染河水、活性污泥、沉积物和土壤中微生物的可培养性分别为0. 001 ～ 0. 1 %、0. 25 %、0. 1 ～ 1 %、0. 1 ～3 %、1 ～15 %、0. 25 %和0. 3 %[2] 。如何进一步对这些实验室内不可培养的微生物资源进行研究开发具有巨大的理论和应用价值。

随着PCR 技术、DNA重组技术和高通量基因组测序技术等的逐渐成熟与完善,以及生物信息学等相关学科的发展,利用分子生物学技术直接研究环境中不可培养微生物基因资源成为可能。1986 年,Olsen GJ 等[3] 提出直接从环境中克隆微生物16S rDNA ,从而开启了以分子生物学方法研究不可培养微生物的大门。1991 年, Schmidt TM 等[4] 通过克隆北中大平洋海水微浮游生物样品中的DNA构建了第一个噬菌体文库。1996年 , DeLong 小组[5] 构建了海水环境DNA的基因组文库,并鉴定了一个从未培养过的古细菌的16S rRNA 基因。1998 年, Handelsman J 等[6]在前人研究的基础上,正式提出了宏基因组(metagenome) 的概念,其定义为“The genomes of the total microbiota found in nature”,即特定生境中全部微生物群遗传物质的总和。

所谓宏基因组学就是利用非培养的分子生物学技术方法对宏基因组进行系统研究,即分析微生物在环境中的基因组集合,研究其群落结构与生态功能等。宏基因组技术流程为: 1) 从特定环境中提取全部DNA ;2)制备一定长度的DNA 片段并连接到合适的载体上;3) 转化宿主菌,构建重组的DNA文库即宏基因组文库;4)

宏基因组文库筛选或者高通量DNA测序分析。宏基因组技术采取非培养方法,从DNA提取入手构建宏基因组文库,直接对环境中大量未能培养的微生物基因资源进行研究，极大的推动了环境微生物生态学的发展,为人类开发利用自然环境中微生物基因资源开辟了新的途径



Handelsman J , et al. Chem Biol.1998 Oct;5(10):R245-9

图 1.1 宏基因组学示意图

宏基因组学的出现对于系统微生物学的研究具有巨大的方法学意义，对于微生物群落的系统研究包括了基因组学、转录组学、蛋白质组学、宏基因组学、代谢组学等技术方法的综合运用。系统微生物学作为系统生物学的一个分支，其主要目标是理解微生物生命现象的整体线路图，研究各个原件（基因、蛋白、生物分子等）怎样一步步构成生命个体、微生物群落及生态系统。与简化的还原论不同，系统生物学强调整合各种生物学相关的手段方法来研究一个微生物细胞或者微生物群落是如何作为一个整体表现出生命现象的[7]。

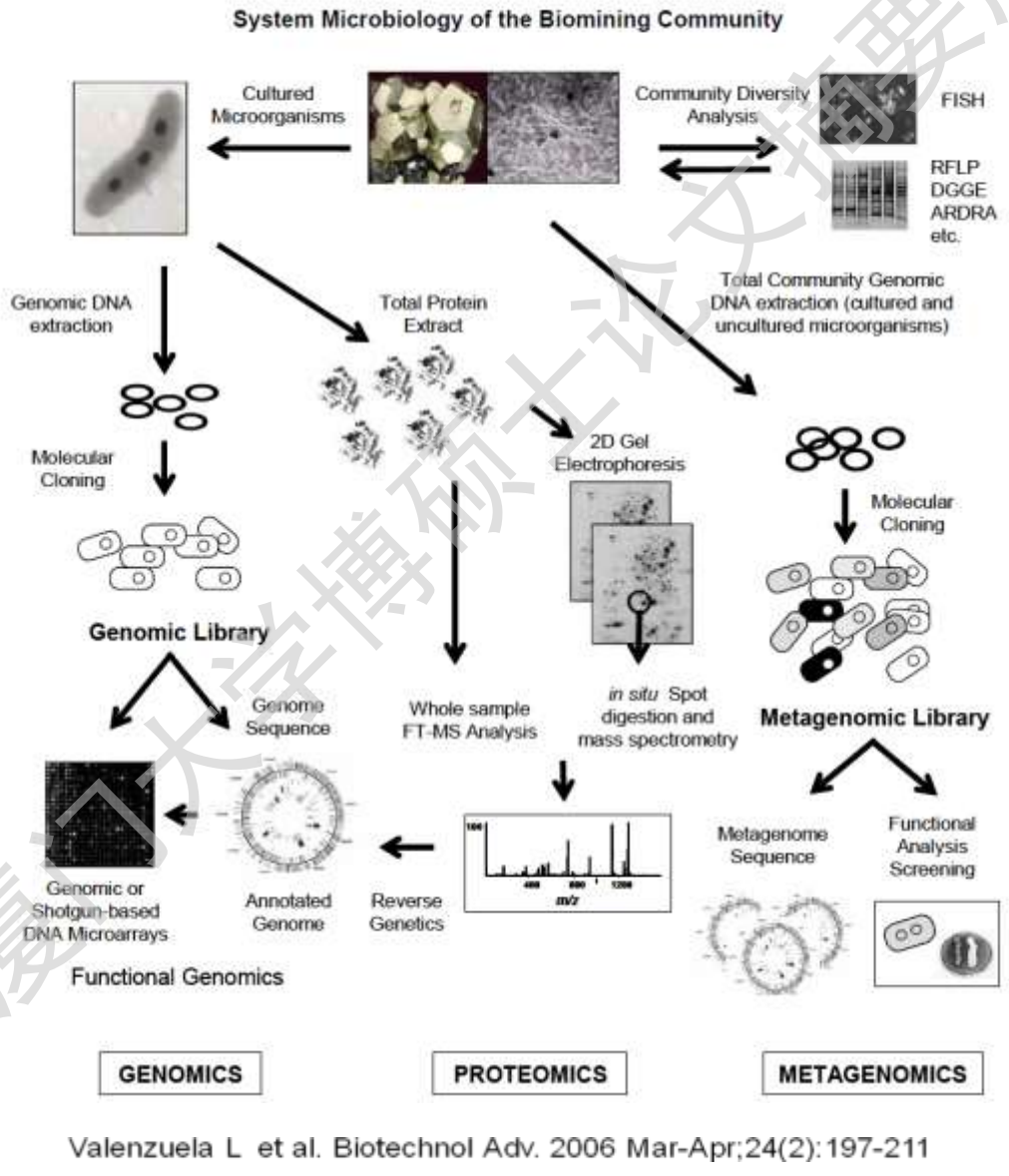


图1.2 系统微生物学示意图

自1998年宏基因组概念提出至今，宏基因组学得到长足的发展，目前研究

主要集中在两方面：一是对不同生境中宏基因组进行高通量测序及生物信息学分析，获得环境微生物生态学数据，从宏基因组学角度探讨复杂微生物系统的组成、进化、代谢过程与环境生态系统之间的关系；二是利用宏基因组文库发现新的功能基因、基因簇及其表达产物，并应用于人类的健康、工业、农业、环保等领域。

对不同生境宏基因组的高通量测序包括了鸟枪法[8-17]和454测序法[18-32]，截止到2008年度，获得的宏基因组数据如表1.1和表1.2。

表1.1鸟枪法对不同生境微生物群落的宏基因组测序分析

Sample	Library size*	Host or vector system used	Average insert size (kbp)	Biodiversity	References
Sargasso Sea	1 985 561	Bst XI linearized pBR322 derivative	2–6	Samples were dominated by genes from *Proteobacteria* (primarily subgroups *Alpha*, *Beta*, and *Gamma*) with moderate contributions from *Firmicutes*, *Cyanobacteria*, and species in the CFB phyla (*Cytophaga*, *Flavobacterium*, and *Bacteroides*). Poor sequencing coverage enabled the assembly of only two near-complete genomes. Here, 1.6 Gbp of unique metagenomic DNA sequences were obtained	Venter *et al.* (2004)
Human feces	36 769	Zero Blunt TOPO PCR cloning	0.5–1.0	Study of uncultured viruses in human feces. The most abundant fecal virus was pepper mild mottle virus	Zhang *et al.* (2006)
Human distal gut	139 521	pHOS2 (*S3*)	2–3	72 bacterial phylotypes and one archaeal phylotype were identified. The bacterial phylotypes were assigned to only two divisions, the *Firmicutes* and the *Actinobacteria*	Gill *et al.* (2006)
Soil	1129 (Bacteria) 527 (Archaea) 919 (Fungi) 4577 (Viruses)	pCR®2.1-TOPO (Bacterial, Archaeal and Fungal) pSMART (viral)	0.5	This is the first study to use sequencing to characterize soil viral communities. Within each of the four microbial groups, data showed minimal taxonomic overlap between sites, suggesting that soil archaea, bacteria, fungi, and viruses are globally as well as locally diverse	Fierer *et al.* (2007)
Acid mine drainage biofilm	103 462	pUC18	3.2	Authors report the reconstruction of near-complete genomes of *Leptospirillum* group II and *Ferroplasma* type II, and partial recovery of three other genomes	Tyson *et al.* (2004)
Chesapeake Bay virioplankton	564	pSMART-HCK	1.3	This report describes the first detailed examination of an estuarine double-stranded DNA viral metagenome. This analysis suggests that dsDNA viruses are likely one of the largest reservoirs of unknown genetic diversity in the biosphere	Bench *et al.* (2007a, b)
Global Ocean	7 697 926	Bst XI linearized pBR322 derivative	2	Authors report a metagenomic study of the marine planktonic microbiota in which surface (mostly marine) water samples were analysed as part of the Sorcerer II Global Ocean Sampling expedition. The resulting 7.7 million sequencing reads from 41 samples provide an unprecedented look at the great diversity and heterogeneity in naturally occurring microbial populations	Rusch *et al.* (2007)
Soil	1 186 200	pJN105/pCF430 & pBeloBAC11	2.7–45	Authors designed a metagenomic analysis to isolate antibiotic resistance genes from six libraries of soil. They identified nine clones expressing resistance to aminoglycoside antibiotics and one expressing tetracycline resistance	Riesenfeld *et al.* (2004) Rondon *et al.* (2000)
Worm lacking a mouth, gut & nephridia	279 157 (3 kb library) 36 095 (35 kb library)	pMCL200 (3 kb library) pCC1FOS™ (35 kb library)	3–35	Metagenomic approach to describe four co-occurring symbionts from the marine oligochaete worm *Olavius algarvensis*. Data revealed that the symbionts are sulphur-oxidizing and sulphate-reducing bacteria, all of which are capable of carbon fixation, thus providing the host with multiple sources of nutrition	Woyke *et al.* (2006)

*Number of reads produced.

Vieites JM et al.FEMS Microbiol Rev. 2009 Jan;33(1):236-55